International Journal of Information Management 36 (2016) 397–402

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Collaborative filtering with facial expressions for online video recommendation

Il Young Choi[a], Myung Geun Oh[b], Jae Kyeong Kim[b,*], Young U. Ryu[c]

[a] School of Dance & Culture Item Factory Center, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Republic of Korea
[b] Department of E-business, College of Business Administration, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-701, Republic of Korea
[c] Jindal School of Management, University of Texas at Dallas, Richardson, TX 75080, United States

ARTICLE INFO

Article history:
Received 22 August 2015
Received in revised form 20 January 2016
Accepted 22 January 2016
Available online 10 February 2016

Keywords:
Online video recommender system
Facial expression
Personalization
Collaborative filtering

ABSTRACT

Online video recommender systems help users find videos suitable for their preferences. However, they have difficulty in identifying dynamic user preferences. In this study, we propose a new recommendation procedure using changes of users' facial expressions captured every moment. Facial expressions portray the users' actual emotions about videos. We can utilize them to discover dynamic user preferences. Further, because the proposed procedure does not rely on historical rating or purchase records, it properly addresses the new user problem, that is, the difficulty in recommending products to users whose past rating or purchase records are not available. To validate the recommendation procedure, we conducted experiments with footwear commercial videos. Experiment results show that the proposed procedure outperforms benchmark systems including a random recommendation, an average rating approach, and a typical collaborative filtering approach for recommendation to both new and existing users. From the results, we conclude that facial expressions are a viable element in recommendation.

1. Introduction

As the Internet cyberspace increases in its size and user population, the frequency of users' interactions with multimedia products, in particular video products, also grows rapidly. According to Video Brewery,[1] there are approximately 100 million Internet users who watch online videos every day and almost 45.4% of Internet users watch more than one video in a month. Consequently, video marketing has become a popular marketing platform. Many online video providers like Dailymotion, YouTube, and Pandora TV make efforts not only to commercialize their video products through an advertising revenue model, but also to increase Website traffic as a means of improving commercialization (Krishnan & Sitaraman, 2013). In particular, video content providers introduce recommender systems helping users find videos relevant to their preference in order to increase Website traffic.

Online video recommendation, however, faces a unique problem. While users watch videos, they often lose their interests in them. These days, most online video clips run under 10 min (Davidson et al., 2010) and it is very difficult to capture and reflect changes of users' interests over a very short time period in profiles built based on users' historical data. Thanks to recent advances in digital image processing technology, it is possible to trace a video viewer's facial expressions in real-time. It has been argued that facial expressions can play a critical role in predicting a user's preference (Ekman & Oster, 1979; Somerville, Fani, & McClure-Tone, 2011). For instance, a user knits the brows while watching brutal scenes. In brief, facial expressions of target users allow predictions of their emotion or degree of concentration. We can utilize such facial expressions to recommend videos to users. Further, we can discover dynamic and changing user preferences while a user watches a video.

In this study, we propose a procedure based on the changes of facial expressions to address variations of user preferences within a short time period. When a user is exposed to a video, his emotion or concentration is captured by a Web cam in the monitor. Then his emotion or degree of concentration is compared with those of existing users, so that we can decide to continue the video-watching session or quickly propose new video products. Though the proposed procedure basically follows the principle of collaborative filtering (CF), we create and maintain a dynamic user profile using the changes of a user's facial expressions every moment consisting of 64 feature values, instead of using a user's previous purchase or rating records.

A notable point is that the proposed approach does not distinguish new users (with no previous purchase or rating records) from

* Corresponding author.
E-mail addresses: choice102@khu.ac.kr (I.Y. Choi), namanic@naver.com (M.G. Oh), jaek@khu.ac.kr (J.K. Kim), ryoung@utdallas.edu (Y.U. Ryu).

[1] http://www.videobrewery.com/blog/18-video-marketing-statistics.

http://dx.doi.org/10.1016/j.ijinfomgt.2016.01.005
0268-4012/

existing users (whose previous purchase or raring records are available). Thus, the proposed procedure does not suffer from the new user problem, that is, the difficulty in recommending products to users whose past purchase or rating records are not available.

We evaluate the proposed procedure through comparison with a random system, an average rating-based system, and a traditional CF-based system as benchmark systems using thirteen footwear advertising videos. Experiment results indicate that the proposed procedure outperforms the benchmark systems in recommendation to both new and existing users. Based on the experiments, we claim that the use of facial expression data is a viable element in recommendation. Further, it can naturally address the new user problem.

2. Related works

2.1. Online video recommender systems

According YouTube statistics,[2] there are more than 1 billion unique monthly visitors who watch videos over 6 billion hours each month. In general, we have seen a rapid increase in demand for online videos. Accordingly, the digital video advertising market value has been estimated to be approximately $4.15 billion in 2013 (Olmstead, Mitchell, Holcomb, & Vogt, 2014). In such a circumstance, online video recommender systems have a great importance in attracting more users. Users are known to revisit a video Website if they are satisfied with the recommended videos (Jin & Su, 2009).

There are relatively a small number of studies on online video recommendation compared with recommendation in other domains. Mei et al. (2007) proposed an early online video recommender system that utilizes multimodal relevance between video data (i.e., textual, visual, and aural information) and click-through data. Xu, Jiang, and Lau (2008) proposed a personal online recommender system based on tracking of a user's attention time. Arapakis et al. (2009) introduced a methodology for integrating facial expression into user profiling to capture the dynamic user preference as well as the static user preference. Hopfgartner and Jose (2010) introduced a semantic user profile construction scheme for a news video recommender system, in which an implicit fingerprint represented by a user's video interests and degree of concerns is utilized for user profiling. Park et al. (2011) proposed a video recommender system based on tag data aggregations. Because their system is a hybrid system that combines collaborative filtering (CF) with content-based filtering (CBF), it was claimed to address the sparsity problem well. Zhao, Yao, and Sun (2013) proposed a video classification and recommendation methodology based on affective analysis of facial expression.

However, most existing studies have drawbacks. They have the new user problem because they all use rating data for a user profile. Further, such rating data are rather static and inadequate to reflect dynamic and changing user preferences when a user watches a video over a short time period. Although users implicitly express their interests through the changes of the facial expressions every moment they watching a video, this information is ignored when their preferences are predicted.

2.2. Facial expression recognition

A user's facial expressions reveal his emotion (Fasel & Luettin, 2003). For instance, the upper lip rises in anger or the cheek rises in happiness (Carroll & Russell, 1997). The recognition of facial expressions can be a basis in predicting users' preferences. Thanks to the development of digital image processing technologies, it is possible to track a face and recognize facial expressions. Accordingly, a variety of studies related to facial expressions have been presented. Previous studies related to facial expressions are broadly classified into those on the recognition of facial action units or muscular activity that produces facial appearance changes and those on facial expression recognition and analysis.

Researchers proposed systems based on a hidden Markov model (Lien, Kanade, Cohn, & Li, 2000). permanent facial features such as brows, eyes, and mouth and transient facial features such as deepening of facial furrows (Tian, Kanade, & Cohn, 2001), dynamic Bayesian network (Tong, Liao, & Ji, 2007), support vector machine and the texture extraction method (Kotsia, Zafeiriou, & Pitas, 2008), 3-D face tracking (Tsalakanidou & Malassiotis, 2010), and support vector machine regression (Savran, Sankur, & Taha Bilge, 2012) for recognition of facial action units. Others researchers proposed systems using brain mapping (Adolphs, Damasio, Tranel, & Damasio, 1996), functional magnetic resonance imaging (Andersen et al., 2001), Gabor wavelets and learning vector quantization (Bashyal & Venayagamoorthy, 2008), local binary patterns (Shan, Gong, & McOwan, 2009), manifold learning methods (Xiao, Zhao, Zhang, & Shi, 2011), and eye tracking (Lischke et al., 2012) for facial expression recognition and analysis. In this study, we extract the values of facial features that are used for dynamic user profile generation and maintenance.

3. Collaborative filtering with facial expression

A typical previous online video recommender system recommends a video to a user based on his historical usage or interaction data such as click-through data. Thus, it suffers from the new user problem: It cannot recommend videos to a new user because the new user's historical data do not exist. The proposed procedure, called facial expression (FE)-based system, follows the basic principle of collaborative filtering (CF), but modifies the data representation phase of the CF-based system. In other words, the FE-based system creates a user profile using dynamic user preferences obtained from changes of the user's facial expressions captured while watching videos instead of his historical usage or rating data. Subsequently, the FE-based system can recommend videos to new users.

The mechanism behind the FE-based system consists of the four steps. In the first step, it collects a user's facial expression data while he watches a video. Also collected is the user's rating on the video after he finishes the video. In the second step, the collected facial expression data are pre-processed and a dynamic user profile is built on each video. Third, the system identifies the user's neighbors who have exhibited similar facial expression changes. This step is essentially from CF. But the fundamental difference is that the typical CF uses historical usage, purchase, rating, or click-through data while the proposed FE-based system uses facial expression data. In the final step, it predicts a rating on a video that a target user is likely to prefer. The details are as follows.

3.1. Step 1: data collection

Facial expressions consist of facial features including a chin, eyes, eyebrows, a nose, an under lip spots, and others (Fasel & Luettin, 2003). The system collects [x, y]-coordinates of a user's facial features every moment the user is watching a video. A video is composed of k clips and facial expressions are generated by q facial features at each clip. Let $D_{i,f}$ be [x, y]-coordinates of user i's facial expressions at clip f in a video. It is defined as follows:

$$D_{i,f} = \{T_{i,1,f}, T_{i,2,f}, \ldots, T_{i,q,f}\} \quad \text{for} f = 1, 2, \ldots, k,$$

[2] https://www.youtube.com/yt/press/statistics.html.